Technical Sciences
Academy of Romania
www.jesi.astr.ro

Journal of Engineering Sciences and Innovation
Volume 2, Issue 2 / 2017, pp. 27-35

C. Computer Engineering, Technology of Information and Communications

Gaze detection / tracking

RAMONA-GEORGIANA VANGHELE[1*], DUMITRU STANOMIR[1]

[1] *Universitatea Politehnica din Bucureşti, Bucureşti, România*

Abstract. Looking in the nearest future, we will see that technology offers to us the ability to monitor and better control our activities. This ensures optimal and safe functionality. If we focus on the automotive field, we can say that things are evolving very quickly. Nowadays, most of the cars are fitted with electronic devices which greatly enhance piloting experience. The present paper focuses only on a small part of this research area, namely the gaze detection, using the properties of images and video sequences such as brightness, contrast, RGB colors representation. Also, it will be used techniques of contour sharpening and selecting elements in the image.

Keywords: gaze detection, gaze tracking, contour sharpening, point of gaze.

1. Introduction

Over the time, a wealth of technologies has been developed that have attempted to estimate the point at which a person is looking at a particular point in time.
Some of these technologies are presented detailed in the first article [1].
One of these methods is named *Gaze Estimation Using the Face Normal Vector* [1] and uses 4 key points (mouth corners and eye corners) to determine the orientation and the rotation of the person's head (face plane normal vector) using a simplified face model.
Another method is focused on pupils detection and is named *Conic Iris Projection* which is based on the elliptical features of the iris and it extracts the normal vector from the projection of the iris in the image [1]. The objective is to identify the contour of the iris in the image area where the eye is located, and to define an ellipse that matches this contour. Matching the ellipse on the contour will be accomplished by modifying ellipse's parameters that will ultimately provide information about the direction in which the eyeball is oriented.

*Correspondence address: ramona.vanghele.etti@gmail.com

Therefore, the two main categories of operations performed by a gaze tracking system are face tracking operations and operations of pupil center detection. The process compound by these operations is performed for each frame from the input sequence, such that the point to which the person is looking is determined for each frame.

The block diagram presented in the first article [1] highlights the operations of one of these systems. These are divided into several main modules used sequentially to calculate the point to which the person is looking.

In another article, the eye tracking system is developed using a single static camera, but the tracking methods are based on 3D models and requires a good resolution for camera for an accurate POG (point of gase) estimation [2]. This means that it is necessary to identify very small details like radius of the cornea, distance from the center of the cornea to the center of the pupil and then to determinate the index of refraction of the aqueous humor [2] to estimate the point to which a person is looking. Also, a good knowledge of structure of human eyeball is required.

In many POG estimation methods like 2-D Mapping-Based Gaze Estimation Technique [3] or Direct 3-D Gaze Estimation Technique [3], the used camera has a great resolution especially for capturing the glint [3] (cornea reflections) in the image on the basis of which is determined the direction of the gaze.

In terms of the equipment used, there are some techniques which collect information from sensors attached at the person skin around the eyes to determine the eyes movements as in a *Electro-Oculography* [4] or from small coils of wire are embedded in a modified contact lens which are inserted into the eye to measure human eye movements too, as in *Sceleral Search Coils* [4] technique.

The other techniques are non-intrusive and use illuminating methods as *Infrared Oculography* [4] or digital image analysis.

Here, it has been chosen to make a software solution for estimating the gaze of a person with the intention of minimizing the costs and discomfort of the person who use this application. This solution is based on the analysis of digital images and it is a part of the appearance-based gaze estimation [4] class. This refers to the fact that the system uses image content to estimate gaze direction by mapping image data to screen coordinates [4]. These coordinates are then associated with degrees in order to indicate the direction of the person head.

It is considered a video sequence in which the person changes the gaze and the face position, obtained by using an usual webcam embedded in a laptop. This sequence will be applied to the entry of the gaze tracking system and will be processed frame by frame.

In this paper, the driver's gaze tracking will be performed to determine if he or she shows signs of tiredness or inattention when is driving the car. The purpose is monitoring the driver vigilance on the road. We wanted to build a simplified model

of the previously described algorithm for stored video sequences and video sequences captured locally using the webcam available on the laptop.
Therefore, the following section describes the the algorithm and the operating principle of the tracking system, then the third section shows the system structure and details each block of its composition. The role of each block is also explained in this section.
In the fourth section, the results of experiments with this system are presented and possible improvements are added in the next section.
Ultimately, some conclusions about the results and the evolution of the technology are outlined.

2. General Presentation

Will be considered a video sequence captured by a web camera available on a laptop. Camera resolution is 120x160 px. The video sequence is compound by consecutive frames which can be regarded as simple RGB images with 120x160 px resolution.
The video sequence is applied as input to the tracking system which is a software program which uses the digital image analysis for finding in each frame the eyes and the nose with the aim of obtaining information for person's head orientation.
RGB images can be decomposed after each component resulting 3 different images corresponding to amount of red, green and blue [5] [6] in the original image.
If is found in an image a great contrast between 2 surfaces, it is known that the brighter surface has a larger amount of the RGB components. The color of the human face varies from beige to brown, which means that in RGB system are used almost equal amounts to represent the color of a human face in an image. Therefore, it can be chosen any component for processing, because if the face is brighter than the background in the image then the amount of any component will be larger in the face area than in the background.
Based on this idea, the algorithm will choose a component, which looks like a greyscale image and will identify the face as a brighter surface on a dark background and the eyes and nose shadow as darker surfaces on a bright background (face become now the background for eyes and nose shadow).
After the identification, the edge of the face will be marked with red straight lines: the algorithm determines the right, left, top, bottom limits of the surface corresponding to face and places this area in a red rectangle whose sides are located at the values indicated by these limits. Also, the algorithm determines the centers of eyes and nose and marks these with red points.
Finally, the middle of the red rectangle is defined as the position of the nose corresponding to the forward position – the face is oriented forward and the attention of the person is maximum – and charts are designed to show the position of the nose and eyes changes relative to the forward position.

3. Implementation

Therefore, it is necessary to define a new scheme (Fig. 1) for the new simplified algorithm and to explain it further.

Gaze tracking algorithm scheme is presented below:

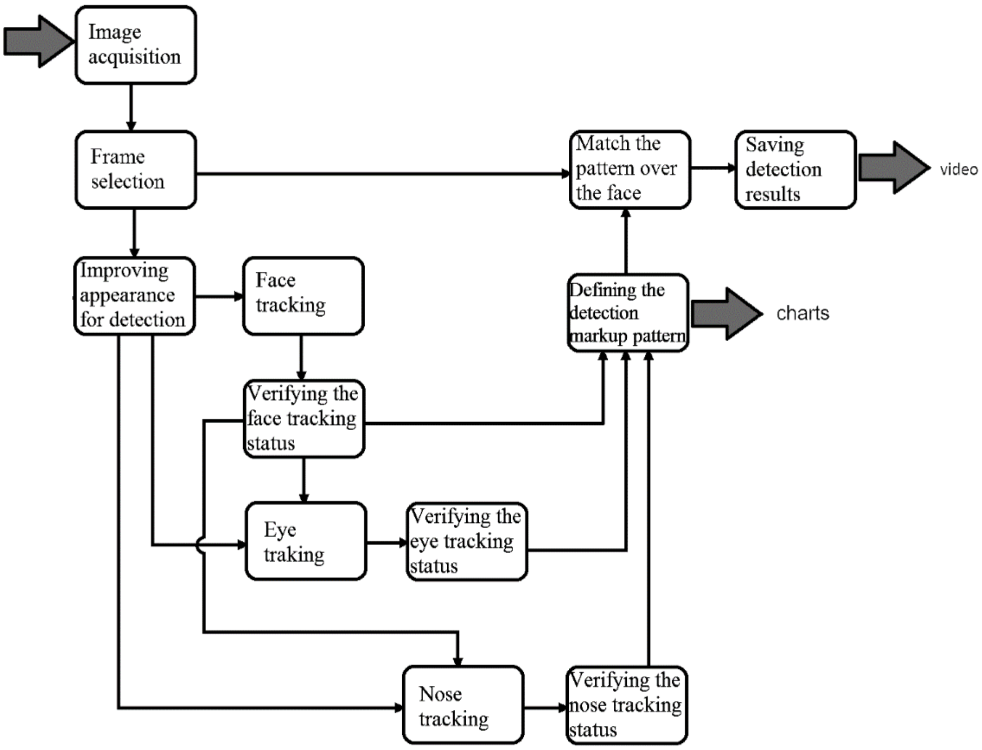

Fig. 1. Gaze tracking algorithm scheme.

The system consists in a Matlab software application developed with the aim to process the video sequence.

Blocks description

Image acquisition can be done directly with the webcam (the situation that corresponds to a real-time application) or it can be considered a recorded video sequence. The second case would correspond to the situation where the driver has a camera installed on his automobile and he or she wants to verify later his/her driving attention.

Frame selection block extracts the current frame from the video sequence. The algorithm is applied to each frame. Here is chosen one of the RGB components of the image to be processed.

Enhancing appearance for detection block uses the Laplace filtering for sharpening and accentuating contours (Fig. 2). This will help to easily identify the edges of the elements in the image. It is necessary to follow this step for an accurate selection of the elements.

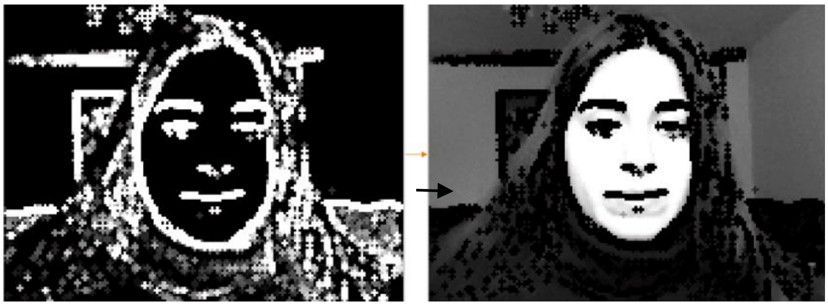

Fig. 2. Enhancing appearance.

Face tracking block, based on contrast between face and background / eyes / shadows creates a binary image representing a mask corresponding to the complexion.
The binary image is created with a chosen threshold so as to these two surfaces will be completely separated. if there are portions with the same brightness in both surfaces, then the face and background cannot be separated and the mask cannnot be built. Because of this the great contrast between face and background is important here. This mask is expanded and a shape is obtained whose contour is actually the contour of the face (Fig. 3). For this shape the boundaries (left-right, top-down) are determined and a red rectangle contour is created to mark the contour of the face.
Eye tracking block is based on an idea similar to the one used for the face; it builds a mask for each eye and determine the corners (outer and inner), then calculate the middle of the eye and mark with a red point.

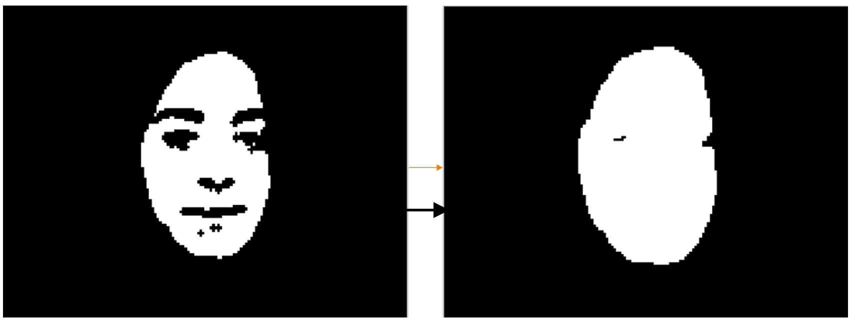

Fig. 3. Determining the shape of the face.

It must specify that this algorithm do not determine with accuracy the pupil, just the middle of the eye. When the light is strong, the shadows and thin dark lines disappear, then the eye can be identified with the iris and the outer and inner corners will become the left-right edges for left eye, or right-left edges for the right eye, of the iris.

Nose tracking block follows the position of the nose through its shadow and marks it with a red point in the top.

Verification blocks indicate whether the detection has occurred successfully, otherwise it breaks the algorithm loop and counts the number of frames whose detection could not be performed. In the final block, the video sequence will be compound only by the frames for which the detection has occurred successfully, because for the rest of the frames it cannot be applied the detection pattern over. In this case, lost frames result from unclear images which is due to fast movements. Ordinary low-resolution video cameras cannot capture this movements, if you want to record fast movements it is necessary to use expensive high-quality digital camcorders. But in this situation ordinary video cameras are enough.

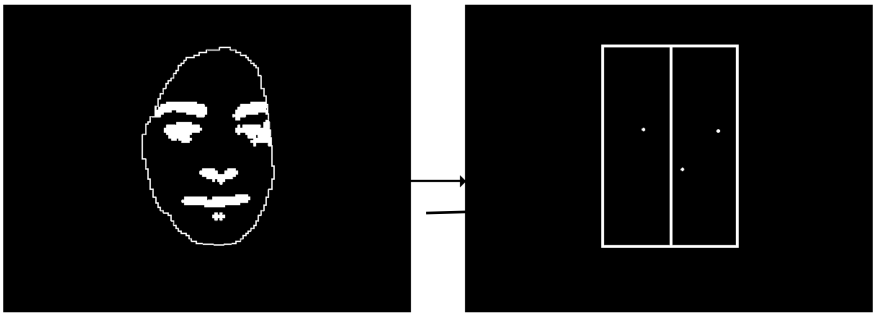

Fig. 4. Detection pattern.

The block corresponding to the **definition of the detection mark pattern** builds the detection markup pattern using the elements obtained from the tracking blocks (face outline, marking points for the eyes and nose). Also, the forward position for the nose and for the eyes result from this block. Using these, the person's attention tracking charts are built.

Saving detection results block overlaps the pattern of detection on the original frames and saves these as a new video sequence with the aim to observe the result of applying this algorithm. Also, a special attention will be given to the number of lost frames.

4. Results

Following the execution of the algorithm, a fairly good accuracy is observed for strong contrast between the face of the person and the background. Otherwise, the algorithm no longer correctly detects face shape and the number of lost frames due to detection failure increases.

The chart corresponding to nose position compared to a forward position allows you to monitor the person's attention on frames: when the degree of nose position relative to the forward position is close to -90 degrees or + 90 degrees, the person has fully turned the face to the left or right and when the degree of nose position relative to the forward position is approaching to 0 degrees, the person has the face in the right position being attentive to driving (Fig. 4).

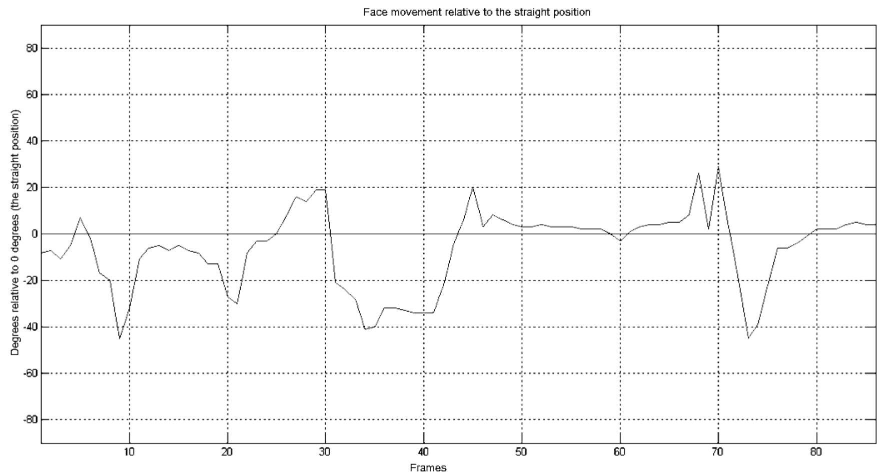

Fig. 4. Face movement chart.

Also, the charts of the positions of each eye allows the person eyes to be monitored in each frame (Fig. 5 and Fig. 6): it can be seen from the charts that both eyes have synchronized movements which confirm the correctness of the detection.
Here, we cannot relate to the forward position defined for the chart that monitors the position of the nose.

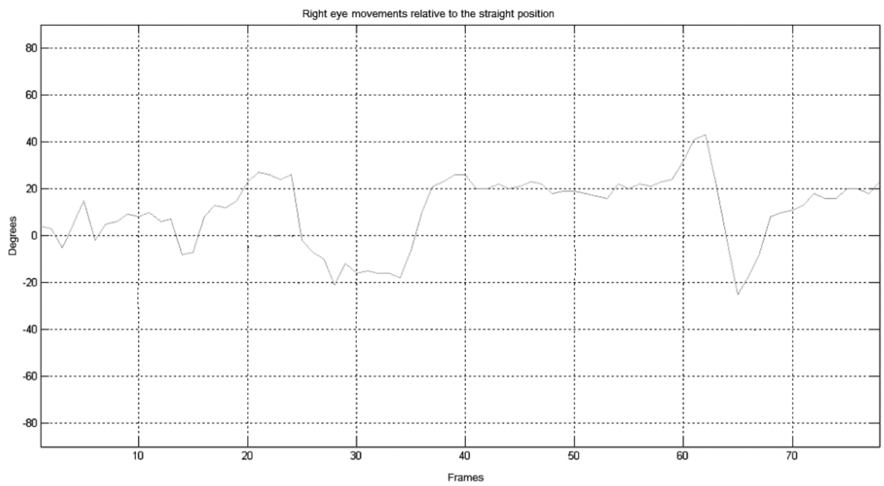

Fig. 5. Right eye movements.

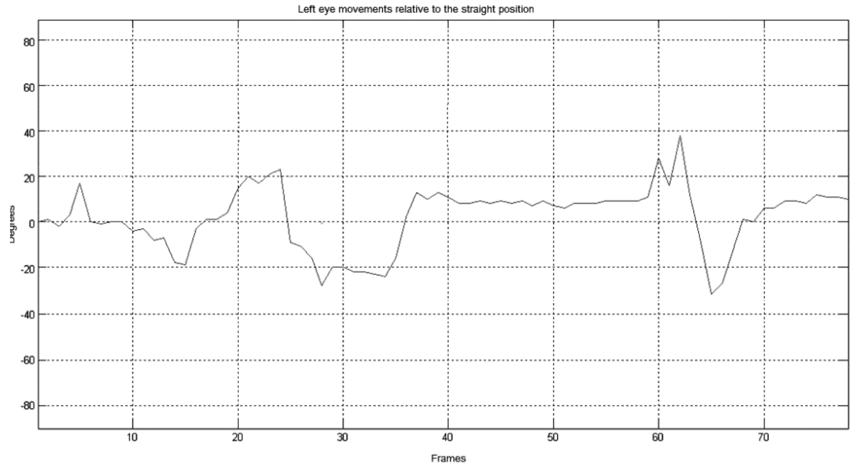

Fig. 6. Left eye movements.

Each eye must have its own reference position to monitor the person's attention. If we fit each eye in a rectangle, then the reference position for maximum attention will be located in the center of each rectangle.
Considering these things, the charts which describe the eyes movements will be similar to the chart which describes the nose movements.

5. Possible improvements

Face detection by color can be a major improvement for this algorithm. Here face detection is made using brightness properties, therefore the face must be in strong contrast with a dark background, otherwise the detection cannot be achieved. In this paper, the image is transformed from RGB to grayscale and only the red part of the image in grayscale was processed. The color determination would involve the use of all 3 RGB components of an image (also in this case it is necessary to customize the color range for each person).
Another improvement is scalability of the image for any resolution, here the resolution used is 120x160 px.
Improve the precision of detecting pupils by looking the pupil in the area between the two corners of the eye will indicates exactly where the person is looking.
And at last, placing limits on charts whose exceeding will means the loss of attention from the forward direction will help in alerting the driver to be careful at the road.
There are many studies in the scientific community, and there are already many detection functions for the face, eyes and the point where a person focuses, but new ideas can appear and improvements can always be made.

6. Conclusion

A person's attention can be monitored with a simple low resolution camera and a software program. It is not mandatory to have the highest precision to say that a person who drives a car is attentive or not when he or she is driving. Also, it is not mandatory to get helpful results to use technologies that can be intrusive and can lead to the discomfort of the person who is driving the vehicle.

We have attempted to achieve in this paper a simpler algorithm based on our current knowledge of image processing.

Using this application, the face, eyes and nose can be detected and followed during the time when the image acquisition takes place and the person's head orientation can be deduced only through digital image analysis.

As results, this system allows you to monitor the person's attention on frames on charts and on a new video sequence.

The nose and eyes positions are compared to reference positions which correspond to forward direction: when the direction deduced by these positions relative to the forward direction is close to -90 degrees or +90 degrees, the person has fully turned the face to the left or right and when the degree of direction deduced by these positions relative to the forward direction is approaching to 0 degrees, the person has the face in the forward being attentive to driving.

Apparently, digital image processing becomes the foundation for the future tracking technologies due to the speed, low cost and non-intrusiveness.

References

[1] I. M. Constantin Carapencea, Real-time gaze tracking with a single camera, *Journal of Information Systems & Operations Management*, p. 13.

[2] B. N. P. L. Craig Hennessey, A Single Camera Eye-Gaze Tracking System with Free Head Motion,p. 8, ETRA 2006, San Diego, California, 27–29 March 2006.

[3] S. M. I. Zhiwei Zhu and Qiang Ji, Novel Eye Gaze Tracking Techniques Under Natural Head Movement, *IEEE Transactions on Biomedical Engineering*, vol. 54, no. 12, DECEMBER 2007.

[4] X. Y. H.R. Chennamma, A survey on eye-gaze tracking techniques, *H.R. Chennamma et.al / Indian Journal of Computer Science and Engineering (IJCSE).*

[5] T. S. T. I. U. P. B. Facultatea de Electronica, Centru de Cercetare pentru Comunicaaţii şi Prelucrări de Semnale, [Online]. Available: http://www.comm.pub.ro/preda/apdsv/laborator_ro.htm. [Accessed 12 07 2017].

[6] T. Ş. T. I. U. P. d. B. Facultatea de Electronică, Centru de Cercetare pentru Comunicaţii şi Prelucrări de Semnale, [Online]. Available: http://www.comm.pub.ro/preda/apdsv/curs_ro.htm. [Accessed 17 06 2017].